

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2002年 8月22日

出 願 番 号

Application Number:

特願2002-242368

[ ST.10/C ]:

[ JP 2002-242368 ]

出 願 人

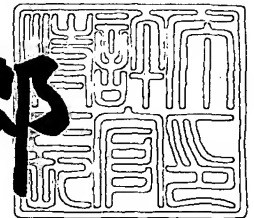
Applicant(s):

株式会社東芝

2003年 2月 7日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3005084

【書類名】 特許願

【整理番号】 13B0250191

【提出日】 平成14年 8月22日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/28

【発明の名称】 機械翻訳システム、機械翻訳方法及び機械翻訳プログラム

【請求項の数】 16

【発明者】

    【住所又は居所】 神奈川県川崎市幸区小向東芝町 1 番地 株式会社東芝  
研究開発センター内

    【氏名】 鈴木 博和

【特許出願人】

    【識別番号】 000003078

    【住所又は居所】 東京都港区芝浦一丁目 1 番 1 号

    【氏名又は名称】 株式会社東芝

【代理人】

    【識別番号】 100076233

    【弁理士】

    【氏名又は名称】 伊藤 進

【手数料の表示】

    【予納台帳番号】 013387

    【納付金額】 21,000円

【提出物件の目録】

    【物件名】 明細書 1

    【物件名】 図面 1

    【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 機械翻訳システム、機械翻訳方法及び機械翻訳プログラム

【特許請求の範囲】

【請求項 1】 原言語から目的言語への翻訳のために前記目的言語による少なくとも 1 つ以上の文書に基づいて構築した目的言語文書データベースを複数管理し、複数の前記目的言語文書データベースに夫々優先度を付すと共に、優先度が最も高い目的言語文書データベースを指定するための通知を行うデータベース管理部と、

前記原言語を目的言語に翻訳して 1 つ以上の訳語候補を発生させる訳語生成部、及び、発生した前記 1 つ以上の訳語候補から前記データベース管理部によって指定された目的言語文書データベースを用いて 1 つの訳語を選択する訳語学習を行う訳語学習部を有する翻訳部とを具備したことを特徴とする機械翻訳システム。

【請求項 2】 前記訳語学習部は、前記データベース管理部によって指定された目的言語文書データベースを用いた訳語選択が不要である単語を示す訳語選択不要単語が前記データベース管理部から指示されて、前記訳語選択不要単語については前記訳語生成部が発生した訳語候補を訳語として確定させることを特徴とする請求項 1 に記載の機械翻訳システム。

【請求項 3】 前記データベース管理部は、ユーザの指示に従って前記訳語選択不要単語を前記目的言語文書データベース毎に設定し、前記訳語学習部に指示することを特徴とする請求項 2 に記載の機械翻訳システム。

【請求項 4】 前記データベース管理部は、前記目的言語文書データベースに設定された前記訳語選択不要単語の数に基づいて前記目的言語文書データベースの優先度を決定することを特徴とする請求項 2 に記載の機械翻訳システム。

【請求項 5】 前記データベース管理部は、前記訳語選択不要単語数及び前記目的言語文書データベースの元となった前記目的言語による少なくとも 1 つ以上の文書と前記原言語との類似度に基づいて、前記目的言語文書データベースの優先度を決定することを特徴とする請求項 2 に記載の機械翻訳システム。

【請求項 6】 前記データベース管理部は、前記訳語選択不要単語数及び前記

類似度に優先させて、ユーザの指定に基づいて前記目的言語文書データベースの優先度を決定することを特徴とする請求項 5 に記載の機械翻訳システム。

【請求項 7】 ユーザ操作に基づいて前記目的言語文書データベースを構築する目的言語文書データベース生成手段を更に具備したことを特徴とする請求項 1 に記載の機械翻訳システム。

【請求項 8】 前記目的言語文書データベース生成手段は、前記目的言語文書データベースに目的言語による文書を追加登録可能であることを特徴とする請求項 7 に記載の機械翻訳システム。

【請求項 9】 前記訳語学習部は、  
前記訳語生成部が生成した訳語候補を格納する訳語候補格納部と、  
ユーザの指示に基づいて訳語選択で使用する評価基準を前記訳語選択部に指示する評価基準選択部と、  
前記訳語選択不要単語を格納する訳語選択不要単語格納部と、  
前記データベース管理部によって指定された目的言語文書データベース、前記評価基準及び前記訳語選択不要単語の情報に基づいて前記訳語候補から最適な訳語を選択する訳語選択部とを具備したことを特徴とする請求項 2 に記載の機械翻訳システム。

【請求項 10】 前記目的言語文書データベース生成手段は、  
前記目的言語による 1 つ以上の文書に対する形態素解析を行う目的言語文書解析部と、

前記目的言語による 1 つ以上の文書から複合語を抽出する複合語抽出部とを具備したことを特徴とする請求項 8 に記載の機械翻訳システム。

【請求項 11】 前記データベース管理部は、前記目的言語文書データベースの優先度の情報を提示する機能を有することを特徴とする請求項 5 に記載の機械翻訳システム。

【請求項 12】 前記データベース管理部は、ユーザの指定による前記目的言語文書データベースの優先度の決定に際して、各目的言語文書データベースの使用による翻訳結果を実際の翻訳前に提示する機能を有することを特徴とする請求項 6 に記載の機械翻訳システム。

【請求項 1 3】 前記訳語学習部は、前記目的言語文書データベースだけでなく、語彙に関する情報を格納する語彙データベースを用いて、前記訳語候補から 1 つの訳語を選択し、

前記データベース管理部は、目的言語文書データベースだけでなく、前記語彙データベースを複数管理し、複数の前記語彙データベースに夫々優先度を付すと共に、優先度が最も高い語彙データベースを指定するための通知を行うことを特徴とする請求項 1 に記載の機械翻訳システム。

【請求項 1 4】 前記データベース管理部は、前記各目的言語文書データベースのデータベース名、前記優先度の情報及び前記訳語選択不要単語の情報を蓄積する管理用データベース蓄積部を有することを特徴とする請求項 1 に記載の機械翻訳システム。

【請求項 1 5】 原言語から目的言語への翻訳のために前記目的言語による少なくとも 1 つ以上の文書に基づいて構築した目的言語文書データベースを複数管理し、複数の前記目的言語文書データベースに夫々優先度を付すと共に、優先度が最も高い目的言語文書データベースを指定するための通知を行う処理と、

前記原言語を目的言語に翻訳して 1 つ以上の訳語候補を発生させる訳語生成処理と、

通知されて指定された目的言語文書データベースを用いて、前記 1 つ以上の訳語候補から 1 つの訳語を選択する訳語学習を行う訳語学習処理とを具備したことを特徴とする機械翻訳方法。

【請求項 1 6】 コンピュータに、

原言語から目的言語への翻訳のために前記目的言語による少なくとも 1 つ以上の文書に基づいて構築した目的言語文書データベースを複数管理し、複数の前記目的言語文書データベースに夫々優先度を付すと共に、優先度が最も高い目的言語文書データベースを指定するための通知を行う処理と、

前記原言語を目的言語に翻訳して 1 つ以上の訳語候補を発生させる訳語生成処理と、

通知されて指定された目的言語文書データベースを用いて、前記 1 つ以上の訳語候補から 1 つの訳語を選択する訳語学習を行う訳語学習処理とを実行させるた

めの機械翻訳プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、訳語学習を自動化した機械翻訳システム、機械翻訳方法及び機械翻訳プログラムに関する。

【0002】

【従来の技術】

近年、コンピュータを利用して、第1言語（原言語）の文章を第2言語（目的言語）の文章に自動的に翻訳する機械翻訳システムが注目されている。機械翻訳システムにおける翻訳処理においては、先ず、入力原文を形態素解析及び構文解析等によって、語（句）等の所定の処理単位に区分する。次に、処理単位毎に翻訳辞書を検索して適用する翻訳規則を決定し、対応する訳語（訳語句）等を決定する。決定した訳語を所定の訳文生成規則に従って結合して、入力原文に対する訳文を得るようになっている。

【0003】

このような機械翻訳において高精度な翻訳を実現するためには、翻訳に最適な辞書を用いることが重要である。種々の原文の翻訳に対応するために、辞書は一般的に同一言語について複数の訳語候補を有している。従って、ユーザは原言語単語に対する複数の訳語候補の中から嗜好や分野、目的等に適した訳語を選択する作業を行う必要がある。

【0004】

ユーザによるこのような選択作業をシステムに学習させることで、以後の訳語選択をユーザに適したものとすることができる。学習作業は、ユーザにとって適切でない訳語が選択されている場合において、その他の訳語候補の中から適切な訳語を選択しなおし、その訳語をシステムに指示することによって行われる。学習作業によって、以降の翻訳ではその訳語が優先的に選択されるようになる。

【0005】

このように、学習作業を通してユーザに適した訳語が選択されるようにする機

能を「訳語学習」という。例えば“computer”の訳語候補としては、「コンピューター」、「コンピュータ」、「計算機」等が存在する。これらの訳語のうちいずれの訳語を選択すればよいかは、ユーザの嗜好や翻訳対象文書の分野、利用する目的等に依存して決まるものである。

## 【 0 0 0 6 】

従来、機械翻訳システムでは、このような学習作業をユーザ自身が行うことによって訳語学習を実現していた。即ち、複数の訳語候補が存在する場合、ユーザにその訳語候補を列挙・提示し、その中からユーザに適切な訳語を選択させシステムに指示し、システムは翻訳時にユーザから指示された訳語を優先的に選択するという方式である。この方式の場合には、適切でない訳語が多ければ多いほどユーザが行う学習作業の回数も多くなり、ユーザに大きな負担をかけてしまう。

## 【 0 0 0 7 】

なお、このような訳語学習については、特開平 9 - 8 1 5 7 2 号公報「翻訳装置及び辞書優先度設定方法」及び特開平 8 - 1 0 1 8 3 6 号公報「機械翻訳の学習方法」等に詳述されている。

## 【 0 0 0 8 】

そこで、ユーザの煩雑な学習作業を不要とする訳語学習方式として、コーパス等の目的言語文書での統計的情報を基に訳語を自動的に決定する方式が提案されている。このような自動訳語学習方式では、目的言語文書をユーザが用意することによってそのユーザに適した自動的な訳語学習が可能になる。

## 【 0 0 0 9 】

なお、これについては、公知文献「目的言語の単言語コーパスを利用した訳語学習方式」（言語処理学会第 8 回年次大会論文集、2002 Vol.1 , pp276-280）及び特開 2 0 0 0 - 2 5 0 9 1 4 号「機械翻訳方法、装置、および機械翻訳プログラムを記録した記録媒体」等に詳述されている。

## 【 0 0 1 0 】

## 【発明が解決しようとする課題】

しかしながら、上述した自動訳語学習方式では、統計的な手法を用いて訳語を決定する。従って、ある特定の単語について不適切な訳語が出力されてしまうこ

ともある。この場合においても、他の目的言語文書を用いて訳語学習を行った場合にはその単語に対しては適切な訳語が出力される可能性もある。しかしながら、複数の目的言語文書の中からいずれの目的言語文書を用いて訳語学習を行った場合に最も適した訳語結果を得ることができるかをユーザが判断することは極めて困難である。例えば、似た内容の目的言語文書が複数存在する場合には、ユーザは各目的言語文書の内容を十分に把握していなければ適切な目的言語文書を選択することはできない。

#### 【 0 0 1 1 】

本発明はかかる問題点に鑑みてなされたものであって、翻訳用データベースに優先順位を付し、ユーザの嗜好、使用目的、使用分野等を反映した自動的な訳語選択を実現することができる機械翻訳システム、機械翻訳方法及び機械翻訳プログラムを提供することを目的とする。

#### 【 0 0 1 2 】

##### 【課題を解決するための手段】

本発明の請求項 1 に係る機械翻訳システムは、原言語から目的言語への翻訳のために前記目的言語による少なくとも 1 つ以上の文書に基づいて構築した目的言語文書データベースを複数管理し、複数の前記目的言語文書データベースに夫々優先度を付すと共に、優先度が最も高い目的言語文書データベースを指定するための通知を行うデータベース管理部と、前記原言語を目的言語に翻訳して 1 つ以上の訳語候補を発生させる訳語生成部、及び、発生した前記 1 つ以上の訳語候補から前記データベース管理部によって指定された目的言語文書データベースを用いて 1 つの訳語を選択する訳語学習を行う訳語学習部を有する翻訳部とを具備したものである。

#### 【 0 0 1 3 】

本発明において、データベース管理部は、目的言語による少なくとも 1 つ以上の文書に基づいて構築した目的言語文書データベースを複数管理し、複数の目的言語文書データベースに夫々優先度を付す。そして、データベース管理部は、優先度が最も高い目的言語文書データベースを指定するための通知を行う。訳語生成部は、原言語を目的言語に翻訳して 1 つ以上の訳語候補を発生させる。発生し



た複数の訳語候補について、訳語学習部は、通知されて指定された目的言語文書データベースを用いて、1つの訳語を選択する訳語学習を行う。

【0014】

なお、装置に係る本発明は方法に係る発明としても成立する。

【0015】

また、装置に係る本発明は、コンピュータに当該発明に相当する処理を実行させるためのプログラムとしても成立する。

【0016】

【発明の実施の形態】

以下、図面を参照して本発明の実施の形態について詳細に説明する。図1は本発明の一実施の形態に係る機械翻訳システムを示すブロック図である。

【0017】

本実施の形態は、原言語から目的言語への機械翻訳において、ユーザが複数の目的言語文書データベースを構築可能にすることにより、訳語学習の精度を向上させることを可能にしている。更に、構築された目的言語文書データベース毎に、訳語の選択を行わない単語である訳語選択不要単語を設定可能にし、訳語選択不要単語の数等に応じてユーザが設定した目的言語文書データベースの優先度を決定することで、有効な目的言語文書データベースの使用を可能にして高精度の翻訳処理を行っている。更に、訳語学習に際して原文に類似した目的言語文書からなる目的言語文書データベースを選択するようになっており、更に一層高精度の翻訳処理が可能である。そして、これらの目的言語文書データベースの選択等を支援する提示を可能にすることにより訳語学習の作業性を著しく向上させるようになっている。

【0018】

なお、訳語選択不要単語については、訳語選択に際して訳語の変更を行わず、訳語学習前に選択されていた訳語を訳語候補として決定する。このような、訳語学習が不要な単語を目的言語文書データベース毎に登録するという考えは、翻訳の際に使用される語彙データベースに対しても応用することができる。即ち、語彙データベース毎に、その語彙データベースを用いて訳して欲しくない単語を設

定しておけば、ある語彙データベースを用いたときに出力されてしまう不適切な訳語を抑制することが可能である。以後、訳語学習が不要な単語及び語彙データベースを用いた訳出が不要な単語を合わせて訳語選択不要単語という。

#### 【 0 0 1 9 】

図 1 において、入力部 1 は、翻訳対象となる文書（原文）の電子データ（以下、単に文書という）を取込んで制御部 2 に出力するようになっている。入力部 1 は、例えば、キーボード、マウス等によって構成されており、ユーザの入力操作に基づく文書の入力が可能である。また、入力部 1 としては、OCR（光学式文字読み取り装置）や、フロッピー（R）ディスク、磁気テープ、磁気ディスク、光ディスク等のコンピュータ可読媒体からの読み込み装置を採用することも可能であり、入力部 1 はこれらの装置によって取込んだ文書を制御部 2 に供給することができるようになっている。また、入力部 1 は、制御部 2 に対して各種コマンド入力を与えることができるようになっている。

#### 【 0 0 2 0 】

出力部 3 は制御部 2 を介して供給された翻訳結果を出力するようになっている。また、出力部 3 は、制御部 3 に制御されて、翻訳結果等を図示しない表示装置の表示画面上に画面表示させることもできるようになっている。また、出力部 3 は翻訳結果を音声出力することもできるようになっている。また更に、出力部 3 は、制御部 2 への各種命令に対する制御部 2 からの応答を表示することもできるようになっている。

#### 【 0 0 2 1 】

なお、出力部 3 としては、各種ディスプレイ等の表示装置だけでなく、印字機等の印刷装置、フロッピー（R）ディスク、磁気テープ、磁気ディスク、光ディスク等のコンピュータ可読媒体への出力装置や、他のメディアに文書を送信する送信装置等を採用することもできる。

#### 【 0 0 2 2 】

翻訳部 4 は制御部 2 からの指示に従って、後述する翻訳辞書部 5 及び語彙データベース蓄積部 8 等を用いて、入力部 1 によって入力された原文に対する翻訳処理を行い、翻訳結果を制御部 2 に出力するようになっている。

## 【 0 0 2 3 】

翻訳辞書部 5 は、図示しない語彙辞書及び翻訳処理に必要な各種規則を格納している。例えば、翻訳辞書部 5 は形態素解析規則、構文・意味解析規則、翻訳の変換規則、翻訳文生成規則等を格納している。語彙データベース蓄積部 8 は、単語・熟語辞書を格納している。

## 【 0 0 2 4 】

図 2 は図 1 中の翻訳部 4 の具体的な構成の一例を示すブロック図である。

## 【 0 0 2 5 】

翻訳部 4 は解析部 4 1、変換部 4 2 及び生成部 4 3 によって構成されている。解析部 4 1 は入力原文に対して翻訳辞書部 5 に格納されている形態素解析規則や語彙データベース蓄積部 8 に格納されている単語・熟語辞書等を用いて形態素解析・辞書引き処理を行い、品詞、活用、訳語等の翻訳処理に必要な情報を求めた後、翻訳辞書部 5 に格納されている構文・意味解析規則を用いて構文・意味解析処理を行う。

## 【 0 0 2 6 】

変換部 4 2 は翻訳辞書部 5 に格納されている変換規則を用いて、構文・意味解析の解析結果から目的言語の構造への変換を行う。生成部 4 3 は変換部 4 2 の出力が与えられ、翻訳辞書部 5 に格納されている生成規則を用いて目的言語の表層文の生成を行い、その結果を制御部 2 に出力するようになっている。

## 【 0 0 2 7 】

生成部 4 3 は、訳語学習部 4 4 を有する。訳語学習部 4 4 は、制御部 2 によって訳語学習の指示がある場合には訳語学習を行い、制御部 2 によって訳語学習の指示がないか、又は訳語学習をしないという指示がある場合には訳語学習を行わない。訳語学習を行う場合には、原言語単語に訳語候補が複数存在するとき、予め準備された目的言語文書データベース蓄積部 9 を用いて、そこに格納されている目的言語文書での統計データや複合語に関する情報等に基づいて、その訳語候補の中から最適な訳語を選択するようになっている。

## 【 0 0 2 8 】

図 3 は図 2 中の訳語学習部 4 4 の具体的な構成の一例を示すブロック図である

## 【 0 0 2 9 】

訳語学習部 4 4 の訳語候補格納部 4 4 1 は、対象となる原言語単語の全ての訳語候補を格納するようになっている。評価基準選択部 4 4 3 は、選択の際に使用する訳語選択評価基準を決定する。訳語選択不要単語格納部 4 4 4 は、訳語学習が不要の単語（訳語選択不要単語）を一時的に格納する。訳語選択部 4 4 2 は、訳語選択不要単語以外の単語に対して訳語候補格納部 4 4 1 に格納されている訳語候補の中から、評価基準選択部 4 4 3 によって選択された評価基準と目的言語文書データベース 4 4 6 とを用いて訳語選択を行う機能を有する。なお、訳語選択部 4 4 2 は、訳語選択不要単語については、ユーザによる選択操作を必要とすることなく、訳語候補を確定させる。

## 【 0 0 3 0 】

訳語学習部 4 4 は、目的言語文書データベースを用いた自動訳語学習機能を実現するだけでなく、その他の訳語学習機能、例えば、ユーザによって最適な訳語が指示され、以降の翻訳ではその訳語を優先的に選択するような訳語学習機能や、翻訳の対象となる文の分野を推定して最適な訳語を自動的に決定する訳語学習機能等を有することも考えられる。この場合には、各訳語学習機能を実現するために必要な構成要素を訳語学習部 4 4 に具備することが可能である。

## 【 0 0 3 1 】

データベース管理部 6 は、ユーザからのデータベースの指示や各データベースに関する情報（後述する優先度等）に基づいて、目的言語文書データベース蓄積部 9 から訳語学習で使用する目的言語文書データベースを選択し、その目的言語文書データベース名を訳語学習部 4 4 に送信するようになっている。

## 【 0 0 3 2 】

目的言語文書データベース作成部 7 は、ユーザが予め用意した任意の目的言語文書から訳語学習部 4 4 で使用する目的言語文書データベースを生成する機能を有している。データベース管理部 6 は、目的言語文書データベース生成部 7 が生成した目的言語文書データベースを目的言語文書データベース蓄積部 9 に蓄積させる。

## 【 0 0 3 3 】

図 4 は図 1 中の目的言語文書データベース生成部 7 の具体的な構成の一例を示すブロック図である。目的言語文書解析部 7 1 は、ユーザが用意した目的言語文書に形態素解析を行い品詞の情報を付加し、その結果を出力する。複合語抽出部 7 2 は、ユーザが用意した目的言語文書中に存在する複合語を抽出し、その結果を出力する。

## 【 0 0 3 4 】

目的言語文書解析部 7 の解析結果及び複合語抽出部 7 2 の抽出結果は、データベース管理部 6 を介して目的言語文書データベース蓄積部 9 に供給される。

## 【 0 0 3 5 】

図 5 は図 1 中の目的言語文書データベース蓄積部 9 の具体的な構成を示すブロック図である。

## 【 0 0 3 6 】

目的言語文書データベース蓄積部 9 の目的言語文書格納部 9 1 は、データベース管理部 6 を介して目的言語文書データベース生成部 7 の目的言語文書解析部 7 1 から品詞情報が付与された目的言語文書が与えられて、この情報を格納する。また、複合語リスト格納部 9 2 は、データベース管理部 6 を介して複合語抽出部 7 2 から目的言語文書中に存在する複合語のリストが与えられて、この情報を格納する。

## 【 0 0 3 7 】

なお、目的言語文書データベース蓄積部 9 は上記構成に限定されるものではなく、実際の利用形態に応じてその他様々な情報（文書の分野や単語の意味・概念情報など）を格納するようにしてもよい。

## 【 0 0 3 8 】

図 6 は図 1 中のデータベース管理部 6 の具体的な構成を示すブロック図である。

## 【 0 0 3 9 】

データベース管理部 6 は、訳語選択不要単語を管理し、翻訳部 4 に訳語選択不要単語に関する情報や優先度等に基づいて決定した使用データベース名を送信す

るようになっている。データベース管理部 6 の翻訳用データベース管理部 6 1 は、語彙データベース蓄積部 8 内の語彙データベースに関するデータを格納する管理用データベース蓄積部 6 3 を管理し、訳語選択不要単語や使用データベースの指示を翻訳部 4 に送信する。

#### 【 0 0 4 0 】

また、目的言語文書データベース管理部 6 2 は、目的言語文書データベース蓄積部 9 内の目的言語文書データベースに関するデータを格納する管理用データベース蓄積部 6 4 を管理し、翻訳部 4 の訳語学習部 4 4 に訳語選択不要単語や使用データベースの指示を送信する。

#### 【 0 0 4 1 】

翻訳用データベース管理部 6 1 及び目的言語文書データベース管理部 6 2 によって、訳語選択不要単語は、対象のデータベース（目的言語文書や語彙データベース等）名と、ユーザによって設定されたデータベースの使用優先度、現在までにそのデータベースを用いて処理された原文データ等と組にされて、管理用データベース蓄積部 6 3, 6 4 に格納される。

#### 【 0 0 4 2 】

図 7 は図 6 中の管理用データベース蓄積部 6 3, 6 4 におけるデータ格納の一例を示す説明図である。

#### 【 0 0 4 3 】

管理用データベース蓄積部 6 3, 6 4 に格納される管理用データベースは、各データベース名毎に、原文データ、訳語選択不要単語及び優先度の情報を有している。管理用データベースは、例えばユーザが初めて各データベースに優先度を設定したときに作成されるようにすればよい。以降はデータベースが追加されたり、訳語選択不要単語を新たに設定したり、翻訳処理を行っていくごとに、管理用データベースの各エントリを更新していくことによって、データベースの最新の使用状況が管理用データベース蓄積部 6 3, 6 4 に格納されるようになっている。

#### 【 0 0 4 4 】

次に、このように構成された実施の形態の動作について図 8 乃至図 1 7 を参照

して説明する。

【 0 0 4 5 】

本実施の形態においては、訳語学習の処理前に目的言語文書データベースを用意するようになっている。目的言語文書データベースは、目的言語文書データベース生成部 7 においてユーザによって予め用意された目的言語文書、あるいは機械翻訳システムが出力する目的言語文書から生成する。

【 0 0 4 6 】

ユーザは、例えば既にハードディスク等に格納されている電子化文書や、インターネット等で入手した電子化文書等を、予め用意された目的言語文書として利用することができる。ユーザが業務や研究等で機械翻訳を利用する場合には、既にユーザのコンピュータ等にはその業務や研究に関する目的言語の文書が格納されていることが多いと考えられ、このような目的言語文書の入手は容易である。

【 0 0 4 7 】

また、一般のユーザであったり専門的な電子化文書を所持していないユーザであっても、近年インターネット等を通して様々な電子化文書を自由に入手することができるようになっており、ユーザが特定の分野に関する電子化文書を用意することも容易である。更に、コンピュータの記憶容量の向上に伴って個人であっても多量の電子化文書を保持することが可能である。ユーザが収集したり予め保持していた電子化専門文書等を用いることにより、そのユーザに特化した目的言語文書データベースを生成することができる。

【 0 0 4 8 】

また、後述する原文書の翻訳結果を、機械翻訳システムが出力する目的言語文書として利用することも可能である。翻訳結果を目的言語文書と見なして再利用することで、そのユーザに適した目的言語文書データベースを構築することができる。

【 0 0 4 9 】

図 8 は目的言語文書データベースを生成するための具体的な目的言語文書データベース生成フローを示すフローチャートである。また、図 9 は目的言語文書データベースを新規生成するための目的言語文書データベース新規生成フローを示

すフローチャートである。図10は目的言語文書データベース生成フロー時に表示されるデータベース生成ダイアログの表示例を示す説明図である。また、図11は目的言語文書データベースの新規生成フロー時に表示されるデータベース新規生成ダイアログの表示例を示す説明図である。

【0050】

まず、入力部1においてユーザが目的言語文書データベースの生成を指示する操作を行うと、制御部2は出力部3によってディスプレイ上にデータベース生成ダイアログを表示する指示を送信する。これにより、出力部3は、図10に示すデータベース生成ダイアログを表示する（ステップS201）。図10のデータベース生成ダイアログには、例えば、既に存在している目的言語文書データベース名を表示して選択を行うための表示欄と、ユーザが用意した目的言語文書を指定するための入力欄が設けられており、また、存在する目的言語文書データベースにデータを追加をするのか、それとも目的言語文書データベースを新規に生成するのかを指定するための選択用のボタン表示も表示されている。

【0051】

制御部2は、ユーザによる選択操作を受け付ける。いま、ユーザがデータベース生成ダイアログ中の「新規生成」ボタンに対する操作によって、目的言語文書データベースの新規作成を指示するものとする。そうすると、制御部2は、処理をステップS202からステップS208、S209に移行して、目的言語文書データベースの新規生成フローを実行する。

【0052】

この場合には、制御部2は、出力部3によってディスプレイ上に図11に示すデータベース新規生成ダイアログを表示させる（ステップS301）。

【0053】

図11のデータベース新規生成ダイアログには、例えば、ユーザが用意した目的言語文書を指定するための入力欄と新規生成する目的言語文書データベース名を指定するための入力欄が表示されている。ユーザは表示されたデータベース名を参照してデータベースを選択すると共に、データベースの元となる目的言語文書名を入力する。制御部2はステップS302、S303において目的言語文書名の



入力及びデータベースの選択が行われたものと判定すると、ステップ S 304 において指定された目的言語文書に対して形態素解析を行う。

## 【 0 0 5 4 】

次に、制御部 2 は、ステップ S 305 において、形態素解析結果によって品詞情報が付加された目的言語文書をステップ S 303 で指定された目的言語文書データベースの目的言語文書を格納する目的言語文書格納部 9 1（図 5 参照）に保存させる。なお、制御部 2 は、ステップ S 306 において、目的言語文書に複合語が存在するものと判定した場合には、目的言語文書中に存在する全ての複合語をステップ S 303 で指定された目的言語文書データベースの複合語を格納する複合語リスト格納部 9 2 に保存させる（ステップ S 307）。

## 【 0 0 5 5 】

最後のステップ S 308 において、制御部 2 は、生成した目的言語文書データベースをデータベース管理部 6 の管理用データベース蓄積部 6 4 に登録しておく。

## 【 0 0 5 6 】

このようにして、ユーザは用意した任意の目的言語文書を既存の目的言語文書データベースに追加するだけでなく、新たに目的言語文書データベースを生成することができるので、全く分野の異なる文書も分野ごとにまとめてデータベース化しておくことが可能である。

## 【 0 0 5 7 】

次に、既存の目的言語文書データベースにデータを追加する場合の動作について説明する。この場合には、ユーザは図 1 0 のデータベース生成ダイアログ中のデータベースを選択し目的言語文書名を入力した後、追加ボタンを操作する。制御部 2 はステップ S 203 において、データベース生成ダイアログによって目的言語文書が指定されたものと判定すると、その文書に対して形態素解析を行う（ステップ S 204）。次に、ステップ S 205 において、解析結果の品詞情報が付加された目的言語文書を、選択されている目的言語文書データベースに追加して目的言語文書格納部 9 1 に格納する（ステップ S 205）。また、制御部 2 は、ステップ S 206 で目的言語文書に複合語が存在するものと判定した場合には、この複合語を複合語リスト格納部 9 2 に追加する（ステップ S 207）。

## 【 0 0 5 8 】

このようにして、ユーザが用意した任意の目的言語文書を、既に存在している目的言語文書データベースに自由に追加することができ、そのユーザの使用に適した目的言語文書データベースの生成が可能である。

## 【 0 0 5 9 】

本実施の形態においては、例えば予め様々な分野の目的言語文書データベースが用意しておくことができ、ユーザの使用分野にあった訳語選択が可能になり、同じ分野でも異なる使用目的ごとに目的言語文書データベースが用意してあれば、その使用目的やユーザの嗜好に適した訳語学習が行うことができる。

## 【 0 0 6 0 】

図 1 2 は制御部 2 におけるユーザとの対話処理の流れの一例を示すフローチャートである。

## 【 0 0 6 1 】

本実施の形態においては、制御部 2 はユーザとの対話処理によって翻訳処理や訳語学習処理を制御するようになっている。即ち、入力部 1 を介した原文の入力が発生すると（ステップ S 601 ）、制御部 2 は出力部 3 に対して原文の表示を指示する（ステップ S 602 ）。次に、ユーザによる翻訳の指示が入力部 1 に与えられると（ステップ S 603 ）、制御部 2 は翻訳実行命令を翻訳部 4 に送信すると同時に、原文のデータを翻訳部 4 に送りステップ S 604 の翻訳処理を開始させる。

## 【 0 0 6 2 】

一方、ユーザによる入力操作が翻訳指示の操作でない場合には、制御部 2 は処理をステップ S 609 に移行して入力操作が目的言語文書データベース生成指示操作であるか否かを判定する。制御部 2 にユーザから目的言語文書データベースの生成指示が与えられると（ステップ S 609 ）、制御部 2 は目的言語文書データベース生成部 7 にデータベース生成命令を送る。これにより、目的言語文書データベース生成部 7 は、ステップ S 610 において図 8 に示す目的言語文書データベース生成処理を実行し、生成された目的言語文書データベースが目的言語文書データベース蓄積部 9 に蓄積される。

## 【 0 0 6 3 】

また、ユーザから処理全体の終了指示が入力部 1 に与えられると（ステップ S 611 ）、制御部 2 は全ての処理を終了する。それ以外の入力に対しては入力に応じた処理を行う（ステップ S 612 ）。

#### 【 0 0 6 4 】

ステップ S 604 の翻訳処理は翻訳部 4 によって実行される。翻訳部 4 は、翻訳辞書部 5 に格納されている知識（形態素解析規則、構文・意味解析規則、変換規則、生成規則等）や語彙データベース蓄積部 8 に格納されている単語・熟語辞書を利用し、目的言語への翻訳処理を進める。こうして、生成部 4 3 において原文の言語に対する訳語が生成される。

#### 【 0 0 6 5 】

訳語学習の指示が与えられていない場合には、翻訳部 4 の生成部 4 3 は訳語学習を行わない。この場合には、生成部 4 3 における翻訳結果が制御部 2 に返され、制御部 2 は翻訳結果を出力部 3 に送信してユーザに提示する（ステップ S 608 ）。

#### 【 0 0 6 6 】

ここで、生成部 4 3 における訳語生成処理に際して、制御部 2 から訳語学習実行命令が与えられているものとする。本実施の形態においては、上述したように、1 つ以上の目的言語文書によって構成された目的言語文書データベースを複数構築することができる。そして、システムによって、あるいはユーザによって、翻訳処理に最適な目的言語文書データベースを選択することによって、訳語学習の効率を向上させるようになっている。

#### 【 0 0 6 7 】

即ち、訳語学習に用いる目的言語文書データベースが目的言語文書データベース蓄積部 9 に複数蓄積されている場合には、ステップ S 606 において、目的言語文書データベースの 1 つが選択されて訳語学習部 4 4 に指示され、ステップ S 613 において使用する評価基準が訳語学習部 4 4 に指示されるようになっている。訳語学習部 4 4 は、選択された目的言語文書データベース及び評価基準に基づいて、原言語単語の複数の訳語候補の中から最適な訳語を選択するための訳語学習処理を行う（ステップ S 607 ）。

## 【0068】

本実施の形態においては、このような自動訳語学習に際して、訳語学習が不要である単語をユーザが自由に設定可能にした目的言語文書データベースを採用するようになっている。

## 【0069】

訳語選択不要単語は、データベース管理部6によって管理されるようになっている。また、データベース管理部6は、翻訳部4の生成部43の訳語学習部44において使用する目的言語文書データベースを指定することができる。即ち、データベース管理部6は、訳語選択不要単語を管理し、翻訳部4に訳語選択不要単語に関する情報や優先度等に基づいて決定した使用データベース名を送信する。

## 【0070】

いま、目的言語文書データベース蓄積部9内に複数の目的言語文書データベースが蓄積されているものとする。また、同様に、語彙データベース蓄積部8内に複数の語彙データベースが蓄積されているものとする。また、目的言語文書データベース蓄積部9に蓄積されている目的言語文書データベースのデータベース名が優先度や訳語選択不要単語等の情報と共に、管理用データベース蓄積部64に既に格納されているものとする。同様に、語彙データベース蓄積部8に蓄積されている語彙データベースのデータベース名が優先度や訳語選択不要単語等の情報と共に、管理用データベース蓄積部63に既に格納されているものとする。

## 【0071】

先ず、図13を参照して目的言語文書データベースを用いた訳語学習について説明する。図13はデータベース管理部6内の目的言語文書データベース管理部62における動作を示すフローチャートである。

## 【0072】

いま、ユーザによって、目的言語文書データベース中の所定の単語について訳語選択不要単語とする旨の指示操作が行われるものとする。この操作に基づいて、制御部2は、目的言語文書データベース管理部62に訳語選択不要単語の指示を与える（ステップS1401）。目的言語文書データベース管理部62は、管理用データベース蓄積部64に保持されている対象の目的言語文書データベースにつ

いて、指示された訳語選択不要単語を追加する（ステップ S 1402）。

#### 【 0 0 7 3 】

例えば図 7 の例では、「データベース A」については、訳語選択不要単語として、単語 W 1，W 2，W 3 が登録されていることが示されている。なお、図 7 では、「データベース A」が、原文データとして文書 1 及び文書 2 を用いて構築されたことが示されている。

#### 【 0 0 7 4 】

管理用データベース蓄積部 6 4 は、各データベースに対する訳語選択不要単語の数や割合がある閾値を越えた場合には優先度の変更を行う（ステップ S 1403）。例えば、管理用データベース蓄積部 6 4 は、訳語選択不要単語の数の大小によって、各データベースの優先度を決定する。図 7 の例では、「データベース A」の訳語選択不要単語数は 3 であり、「データベース B」の訳語選択不要単語数は 2 であり、「データベース C」の訳語選択不要単語数は 1 である。仮に、目的言語文書データベース蓄積部 9 にこれらの 3 つのデータベースのみが蓄積されているものとする、データベース A～C の優先度は、夫々 3，2，1 である。優先度は値が少ないほど優先度が高いことを示している。

#### 【 0 0 7 5 】

なお、図 7 の例では、訳語選択不要単語の数の大小のみによって優先度を決定したが、訳語選択不要単語の数が一定数以上になったことによって優先度を決定してもよく、また、格納されている訳語選択不要単語が原文中に存在する割合に応じて優先度を決定してもよい。即ち、訳語選択不要単語については訳語学習が行われないので、訳語選択不要単語を多く含むデータベースを用いる利点は比較的小さいものと判断して、その目的言語文書データベースの優先度を減らし、他のデータベースの優先度を上げる等の変更を行う。この際、優先度の変更された旨をユーザに通知してユーザの判断を仰ぐような手段を具備していてもよい。

#### 【 0 0 7 6 】

これにより、訳語学習が不要の単語を多く含む目的言語文書データベースを無理して使いつづける事がなくなり、自動的な優先度変更により柔軟なデータベース選択が可能となる。優先度の変更方法については既存の技術を応用することが

可能であり、種々の方法が考えられる。

【 0 0 7 7 】

更に、本実施の形態においては、優先度を用いてデータベースを選択するだけでなく、データベースの元となった原文データと入力原文との類似度によっても、訳語学習に用いるデータベースを選択する。

【 0 0 7 8 】

原文データが入力部 1 によって入力され、制御部 2 を介して翻訳部 4 に与えられて翻訳処理が開始されると（ステップ S 1404）、データベース管理部 6 の目的言語文書データベース管理部 6 2 は、管理用データベース蓄積部 6 4 に原文データが格納されているか否かを判定する（ステップ S 1405）。原文データが格納されている場合には、目的言語文書データベース管理部 6 2 は、格納されている原文データと入力原文との類似度を検証する。類似度の検証方法は公知になっている様々な方法を利用可能である。

【 0 0 7 9 】

例えば、類似度の検証方法としては、文書に含まれる単語を次元とするベクトルとして文書を表現する「ベクトル空間モデル」を用いて類似度を形成する手法が広く知られている。なお、これについては、公知文献「文ベクトル集合モデルに基づく文書類尺度の評価」（城塚 晋也・北内 啓（NTTデータ 技術開発本部））（情報研報Vol.2002 No.41 p159-164）等がある。

【 0 0 8 0 】

目的言語文書データベース管理部 6 2 は、入力原文と類似している原文データが存在する場合には（ステップ S 1407）、類似していると判定した原文データに対応する目的言語文書データベースを訳語学習に用いるデータベースとして選択する（ステップ S 1408）。

【 0 0 8 1 】

これにより、以前に訳語学習を行った履歴を用いて使用すべき目的言語文書データベースが決定されることになり、ユーザの使用に適したデータベースの自動的な選択が可能となる。

【 0 0 8 2 】

目的言語文書データベース管理部 6 2 は、原文データが格納されていないか又は類似した原文データが存在しない場合には、例えば訳語選択不要単語数の大小から求めた優先度が最も高い目的言語文書データベースを訳語学習に用いるデータベースとして選択する（ステップ S 1409）。

## 【 0 0 8 3 】

こうして、目的言語文書データベース管理部 6 2 は、ステップ S 1408 又はステップ S 1409 において選択した目的言語文書データベースをシステムによるデータベース選択結果とする。更に、本実施の形態においては、システムによるデータベースの選択だけでなく、ユーザによるデータベースの選択も可能になっている。

## 【 0 0 8 4 】

次のステップ S 1409 における処理 A は、データベース管理部 6 とユーザとの対話処理によってデータベースを選択する処理である。図 1 4 はこの場合の処理の流れの一例を示すフローチャートである。

## 【 0 0 8 5 】

図 1 4 のステップ S 1301 では、データベース管理部 6 はユーザからのデータベースの指示があるか否かを判定している。なお、ユーザは、入力部 1 を介して訳語学習に用いる目的言語文書データベースを指定することができる。データベース管理部 6 は、制御部 2 を介してユーザが指定したデータベース名（ユーザデータベース指示）を受信する。データベース管理部 6 は、ユーザーによる使用データベースの指定がある場合には、指定された目的言語文書データベースを優先的に選択する。なお、この場合には、データベース管理部 6 がシステムによるデータベース選択結果として選択した目的言語文書データベースがある場合には、そのデータベース名を制御部 2 を介して出力部 3 に出力することで、ユーザにその使用を提案する機能を有していてもよい（ステップ S 1306）。

## 【 0 0 8 6 】

ユーザによっては目的言語文書データベースの内容が詳しく分からないことから優先度を設定することができないこともある。そこで、本実施の形態においては、データベース管理部 6 は、各目的言語文書データベースを利用したときに訳

語学習結果がどのようなになるのかを比較・表示する機能を備えている。

【 0 0 8 7 】

即ち、ユーザによって翻訳結果を参照・比較する命令が入力部 1 を介して入力されると（ステップ S 1302）、制御部 2 は選択対象の全てのデータベース名を翻訳部 4 に送信するようにデータベース管理部 6 に命令する。この命令を受けたデータベース管理部 6 は、選択対象データベース名を翻訳部 4 に送信すると共に、各データベース毎に訳語選択不要単語の情報も翻訳部 4 に送信する。

【 0 0 8 8 】

翻訳部 4 は受信したデータベース名のデータベースにアクセスし、各データベース毎に、そのデータベース及び訳語選択不要単語を用いた翻訳結果を出力部 3 に送信する。出力部 3 は、各データベースを使用したときの翻訳結果を例えば画面表示によってユーザに提示する（ステップ S 1303）。

【 0 0 8 9 】

これにより、ユーザはこれらの翻訳結果を比較することができ、どのデータベースを使用すればよいかを容易に決定できるようになる。ユーザ選択によるデータベースは、ステップ S 1304において選択され、ステップ S 1307において、訳語選択不要単語と共に翻訳部 4 内の生成部 4 3 内の訳語学習部 4 4 の訳語選択不要単語格納部 4 4 4 に送信される。また、ステップ S 1408又はステップ S 1409によるシステム選択のデータベースは、ステップ S 1305において選択され、ステップ S 1307において、訳語選択不要単語と共に訳語選択不要単語格納部 4 4 4 に送信される。

【 0 0 9 0 】

なお、制御部 2 はユーザによる翻訳結果参照の指示がない場合は、データベース管理部 6 によって選択されているデータベースを選択する（ステップ S 1305）。

【 0 0 9 1 】

こうして、データベース管理部 6 が原文データとの類似度及び優先度によって選択したデータベース又はユーザ選択に基づくデータベースが、訳語学習のために選択される。これにより、ユーザにとって最適な目的言語文書データベースの



選択が可能となる。

#### 【0092】

翻訳部4は、データベース管理部6によってユーザ選択又はシステム選択に基づくデータベースが指示されると、この指示に従って訳語学習を実行する。図15は訳語学習の具体的なフローを示すフローチャートである。

#### 【0093】

いま、入力原文が入力部1を介して入力され、この原文データが翻訳部4の解析部41、変換部42及び生成部43によって、順次機械翻訳処理されるものとする。そして、翻訳結果として、原言語単語について複数の訳語候補が存在するものとする。この場合には、訳語学習部44は、訳語学習によって訳語を決定する。

#### 【0094】

即ち、先ず、図15のステップS401において、制御部2は、目的言語文書データベースが存在するか否かを判断する。目的言語文書データベースが存在していないのであれば訳語学習は行わず、生成部43は元々選択していた訳語（デフォルト訳語）を訳語候補として選択する（ステップS409）。次のステップS402では、目的言語文書データベースを用いた訳語選択を行わないことを示す命令が発せられたか否かを判定する。この命令が発せられた場合にも、訳語学習は行わず、デフォルト訳語を訳語候補として選択する（ステップS409）。

#### 【0095】

複数の目的言語文書データベースが目的言語文書データベース蓄積部9に蓄積されている場合には、データベース管理部6によって、最適な目的言語文書データベースが選択されて、データベース名が翻訳部4に指示されると共に、訳語選択不要単語が翻訳部4に送信される。

#### 【0096】

データベース管理部6から送信された訳語選択不要単語（図3の訳語選択不要単語445）は、訳語学習部44内の訳語選択不要単語格納部444に格納される。訳語選択部442は、ステップS403において、データベース管理部6によって指示された目的言語文書データベース（図3の目的言語文書データベース4

4 6) を選択する。一方、訳語選択対象の単語の各訳語候補は、生成部 4 3 によって、訳語候補格納部 4 4 1 に格納されている。訳語候補格納部 4 4 1 に格納されている訳語候補は、訳語選択部 4 4 2 に渡される (ステップ S 404)。

#### 【 0 0 9 7 】

ステップ S 405では、訳語選択部 4 4 2 は目的言語文書データベース蓄積部 9 内の目的言語文書格納部 9 1 の選択された目的言語文書データベースにアクセスし、目的言語文書に関する情報を入手する。訳語選択部 4 4 2 は、ステップ S 406において、選択された目的言語文書データベースの複合語リスト格納部 9 2 にアクセスし、目的言語文書中に存在する複合語のリストを入手する。訳語選択部 4 4 2 は、入手した複合語に関する情報や目的言語文書内での共起の強度・出現頻度等に基づいて、最適な訳語を決定する (ステップ S 407)。

#### 【 0 0 9 8 】

図 1 6 は図 1 5 のステップ S 407における訳語選択処理を具体的に示すフローチャートである。

#### 【 0 0 9 9 】

先ず、訳語選択部 4 4 2 は、ユーザによって指定された訳語学習不要単語をチェックする (ステップ S 506)。もし訳語学習不要単語が存在している場合には、その訳語の訳語学習前の訳語を信頼性の高い訳語として扱う (ステップ S 507)。即ち、訳語選択部 4 4 2 は、訳語候補のうち訳語学習不要単語と同一の訳語については、その訳語を確定したものとして、確定訳語リストに追加する。

#### 【 0 1 0 0 】

生成部 4 3 において訳文を生成する際、名詞が 2 語以上続けて並んでいる場合 (ステップ S 501) には、その部分は複合語である可能性が高い。そのとき各名詞のいずれか一方、又は両方に複数の訳語候補が存在する場合には、各訳語候補の組み合わせを求め (ステップ S 502)、使用する目的言語文書データベースの複合語リスト格納部 9 2 に格納されている複合語リストを参照し、そこで用いられている訳語を優先して選択する (ステップ S 503)。

#### 【 0 1 0 1 】

これらの処理によって不確定な訳語は減少し、全体として訳語選択の精度が向

上する。また複合語は分野に大きく依存していると考えられるので、複合語の情報を利用することで、よりユーザの嗜好・使用分野・目的に適した訳語選択が可能となる。

#### 【0102】

例えば英語の“computer magazine”を日本語に訳す場合、「コンピュータマガジン」と訳すか「コンピュータ雑誌」と訳すかはユーザの嗜好等に関わる問題であり、翻訳の仕方はユーザによって、また、同一ユーザでも使用する分野・目的によって異なる。しかしユーザの指定によって生成された目的言語文書データベース中にいずれの用法の単語が用いられているかに応じて確定訳語は決定される。即ち、ユーザの指定による目的言語文書はユーザの嗜好や使用分野・目的が反映されたものであり、生成された目的言語文書データベースの複合語リストにもユーザの嗜好等が反映されている。従って、訳語学習処理において複合語の情報を利用することで、高精度の翻訳が可能となる。

#### 【0103】

訳語選択部442は、ステップS504において、意味規則で決定した訳語や訳語候補が一つしかないもの、複合語リストによって決定した訳語を、信頼度が高い訳語である確定訳語として、確定訳語リストに登録する。なお、確定訳語に意味規則で決定した訳語を追加したのは、意味規則によって決定した訳語が状況に強く依存して決定された訳語であるので、信頼度が高いと思われるからである。

#### 【0104】

次に、訳語選択部442は、ステップS505において、評価基準選択部443によって指示された訳語選択評価基準によって訳語選択を行う。訳語選択の際には、様々な要因（出現頻度、相互情報量等）を考慮した複数の訳語選択評価基準を用意することも考えられる。評価基準選択部443は、このように複数の訳語選択評価基準が存在している場合に有効であり、ユーザによる指示を受けて使用する訳語選択評価基準を決定したり、自動的に使用する訳語選択評価基準を決定する機能を有していてもよい。

#### 【0105】

例えば、評価基準選択部443によって選択された訳語選択評価基準が「出現

頻度優先」であった場合には、目的言語文書データベース蓄積部 9 内の目的言語文書格納部 9 1 に格納されている品詞情報付きの目的言語文書での出現頻度が高いものを優先して訳語選択を行う。また、訳語選択評価基準が、「共起強度優先」である場合には、同様にして各原言語単語の訳語候補同士の共起の強度が大きいものを優先して訳語選択を行う。

## 【 0 1 0 6 】

これらの評価基準は一例であり、他の評価基準を採用してもよいことは明らかである。この場合には、目的言語文書データベース蓄積部 9 に品詞情報付きの目的言語文書と複合語リスト以外の情報とを格納しておき、これらを利用することも可能である。

## 【 0 1 0 7 】

訳語選択部 4 4 2 は、データベース管理部 6 によって選択された目的言語文書データベース 4 4 6 内で、訳語候補格納部 4 4 1 に格納されている各訳語候補を検索・カウントし、指示されている評価基準、及び訳語選択不要単語に基づいて訳語を決定し、生成部 4 3 への出力を行う。生成部 4 3 は訳語学習前の訳語を訳語学習後の訳語に置換して制御部 2 に送る。制御部 2 は翻訳結果を出力部 3 に送信して一連の翻訳処理を終了する。

## 【 0 1 0 8 】

なお、訳語選択不要単語については、生成部 4 3 で始めに生成された元々の訳語を訳語選択部 4 4 2 は確定訳語とするので（ステップ S 507）、訳語学習によって訳語が変更されることはない。

## 【 0 1 0 9 】

次に、図 1 5 のステップ S 408において、訳語学習部 4 4 は、訳語選択部 4 4 2 で選択された訳語を翻訳部 4 の生成部 4 3 に出力して訳語学習を終了する。翻訳部 4 の生成部 4 3 は、入力された訳語を基に訳文を生成し、その結果を出力部 3 に渡して訳語学習された翻訳結果を出力させる。

## 【 0 1 1 0 】

このように、本実施の形態においては、ユーザの嗜好や分野に合わせた目的言語文書データベースを用いて訳語学習を自動的に行うことにより、ユーザ自身に

よる学習作業を軽減しユーザのニーズに合わせた訳出を可能とする。

#### 【0111】

図13のステップS1413において翻訳処理が終了すると、目的言語文書データベース管理部62は、管理用データベース蓄積部64に原文データを格納する（ステップS1414）。次のステップS1415では、新しい目的言語文書データベースがデータベース管理部6に追加されたか否かが判断される。新たな目的言語文書データベースが追加された場合には、データベースの優先度を再設定して管理用データベース蓄積部64の内容を更新する。なお、再び訳語選択不要単語が指定された場合には、ステップS1401以下の動作が繰返される。

#### 【0112】

図13においては、目的言語文書データベース管理部62における動作を説明した。図13の動作を翻訳用データベース管理部61においても実行させることにより、柔軟な辞書データベースの自動選択が可能になる。

#### 【0113】

図17は翻訳用データベース管理部61での動作の一例を示すフローチャートである。図17において図13と同一ステップには同一符号を付して説明を省略する。図17のフローにおいては、ステップS1307に代えてステップS1308を採用した点が図13のフローと異なる。

#### 【0114】

図13のステップS1307では、訳語選択不要単語を訳語学習部44の訳語選択不要単語格納部444に送信している。これにより、訳語選択部442は、指示された訳語選択不要単語については訳語候補をそのまま確定して変更しないようになっている。これに対し、図17のステップS1308においては、指示された訳語選択不要単語の訳出に関しては現在選択されている語彙データベースを使用することなく、別の語彙データベースを優先度に準じて使用させるようになっている。

#### 【0115】

例えば、現在選択されている語彙データベースを「データベースa」とし、その他の語彙データベース名を「データベースb」、「データベースc」とする。

また、設定されている優先度を  $b > a > c$  とする。データベース a における訳語選択不要単語を W としたとき、W はデータベース a を用いて訳出せずに、その他の語彙データベースで優先度の高い方を使用して訳出する。上記の例では「データベース b」を使用して訳出することになる。

## 【 0 1 1 6 】

もし「データベース b」の訳語選択不要単語にも W が登録されていたら次の優先度をもつ「データベース c」を用いて訳出する。こうして訳語が決定するまで上記の操作を繰り返す。使用するデータベースが見つからなかった場合は、優先度が最も大きいデータベースを用いて訳出する。

## 【 0 1 1 7 】

このように、翻訳用データベース管理部 6 1 によって語彙データベースの使用優先度を自動的に変更可能にするだけでなく、単語ごとに異なる語彙データベースを用いて訳出することが可能になる。

## 【 0 1 1 8 】

このように、本実施の形態においては、ユーザが収集した特定分野の目的言語文書を用いて複数の目的言語文書データベースを構築することができる。そして、ユーザによって、各目的言語文書データベース毎に、訳語選択不要単語を設定可能にして、自動訳語学習による不必要な訳語変更を抑制している。更に、ユーザが設定した目的言語文書データベースの優先度を、訳語選択不要単語の状況に応じて自動的に変更可能にし、優先度の情報を用いることで、目的言語文書データベースの有効な選択を可能にしている。また、入力原文と目的言語文書データベースの基となった原文データとの類似判定によって、以前使用した目的言語文書データベースを選択されるようにして、翻訳精度を向上させることができる。更に、各目的言語文書データベースを用いた場合の訳語学習結果を比較表示させることにより、ユーザの目的言語文書データベースの選択作業を容易にすることができる。

## 【 0 1 1 9 】

なお本発明は、上記実施の形態に限定されるものではなく、その趣旨を逸脱しない範囲で適宜変形して実施することができる。

【 0 1 2 0 】

【発明の効果】

以上説明したように本発明によれば、翻訳用データベースに優先順位を付し、ユーザの嗜好、使用目的、使用分野等を反映した自動的な訳語選択を実現することができるという効果を有する。

【図面の簡単な説明】

【図 1】

本発明に係る機械翻訳システムの一実施の形態を示すブロック図。

【図 2】

図 1 中の翻訳部 4 の具体的な構成の一例を示すブロック図。

【図 3】

図 2 中の訳語学習部 4 4 の具体的な構成をの一例を示すブロック図。

【図 4】

図 1 中の目的言語文書データベース生成部 7 の具体的な構成の一例を示すブロック図。

【図 5】

図 1 中の目的言語文書データベース蓄積部 9 の具体的な構成を示すブロック図。

【図 6】

図 1 中のデータベース管理部 6 の具体的な構成を示すブロック図。

【図 7】

図 6 中の管理用データベース蓄積部 6 3, 6 4 におけるデータ格納の一例を示す説明図。

【図 8】

目的言語文書データベースを生成する具体例の目的言語文書データベース生成フローを示すフローチャート。

【図 9】

目的言語文書データベースを新規生成するための目的言語文書データベース新規生成フローを示すフローチャート。

【図 1 0】

目的言語文書データベース生成フロー時に表示されるデータベース生成ダイアログの表示例を示す説明図。

【図 1 1】

目的言語文書データベースの新規生成フロー時に表示されるデータベース新規生成ダイアログの表示例を示す説明図。

【図 1 2】

制御部 2 におけるユーザとの対話処理の流れの一例を示すフローチャート。

【図 1 3】

データベース管理部 6 内の目的言語文書データベース管理部 6 2 における動作を示すフローチャート。

【図 1 4】

データベースの選択処理を説明するためのフローチャート。

【図 1 5】

訳語学習の具体的な処理を示すフローチャート。

【図 1 6】

図 1 4 のステップ S 407 における訳語選択処理を具体的に示すフローチャート。

【図 1 7】

翻訳用データベース管理部 6 1 での動作の一例を示すフローチャート。

【符号の説明】

1 …入力部、 2 …制御部、 3 …出力部、 4 …翻訳部、 5 …翻訳辞書部、 6 …データベース管理部、 7 …目的言語文書データベース生成部、 8 …語彙データベース蓄積部、 9 …目的言語文書データベース蓄積部。

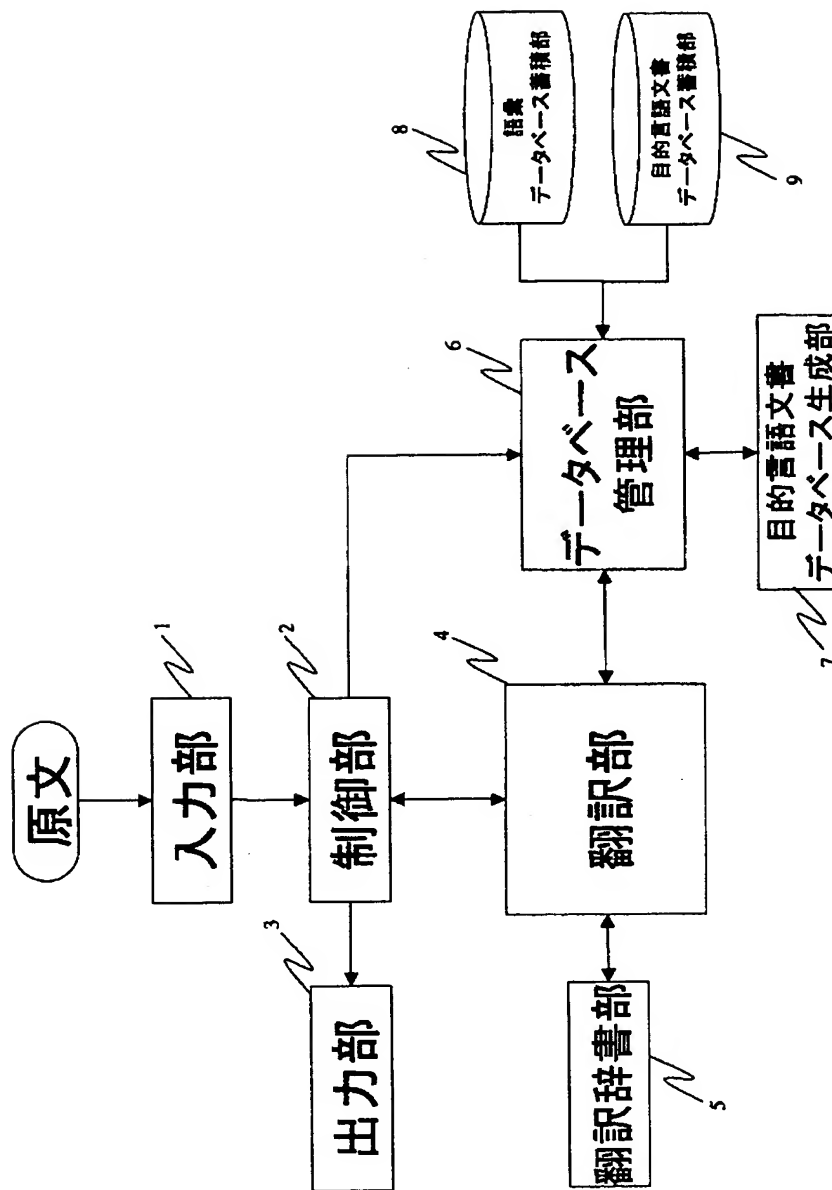
代理人      弁理士      伊 藤      進



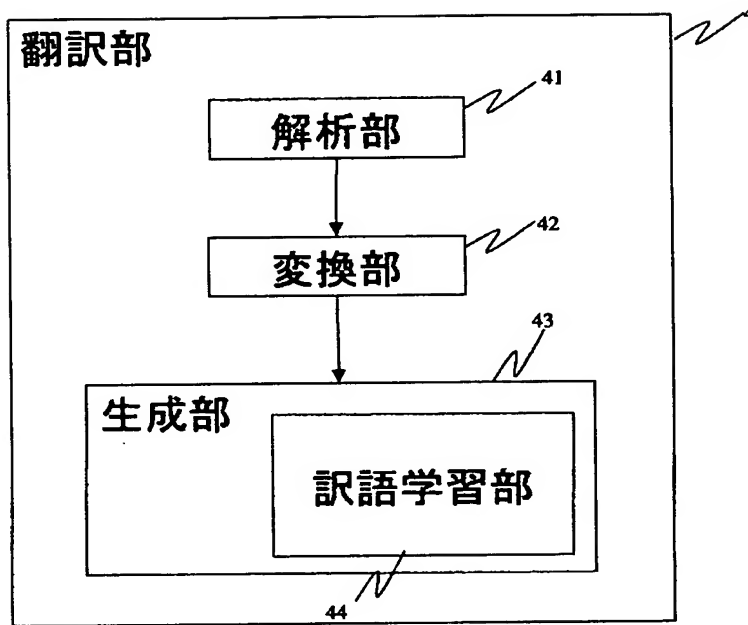
【書類名】

図面

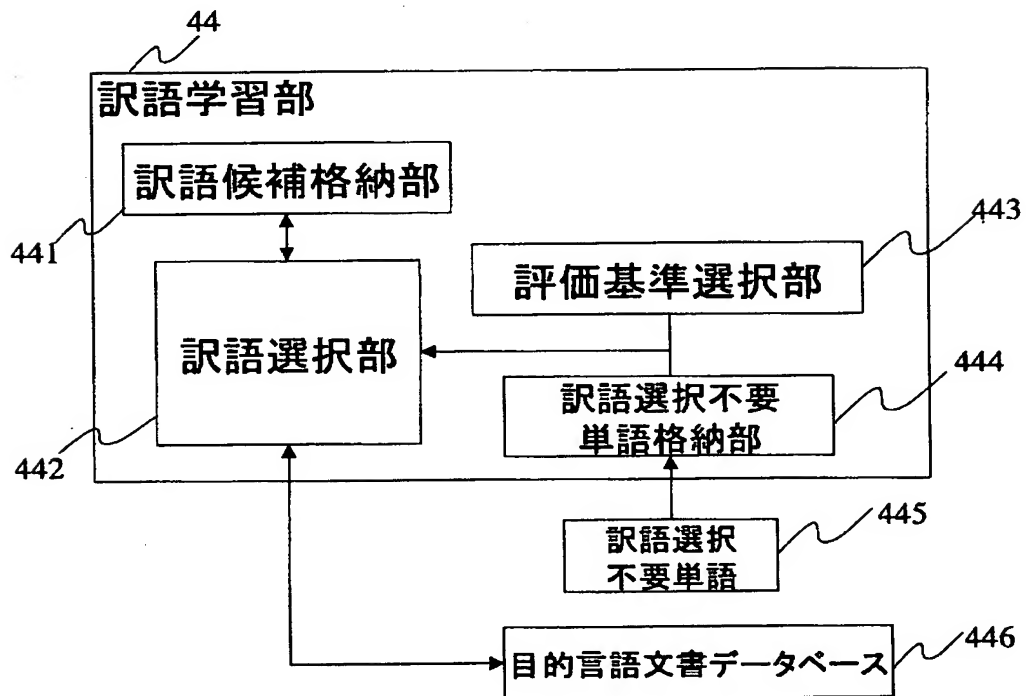
【図 1】



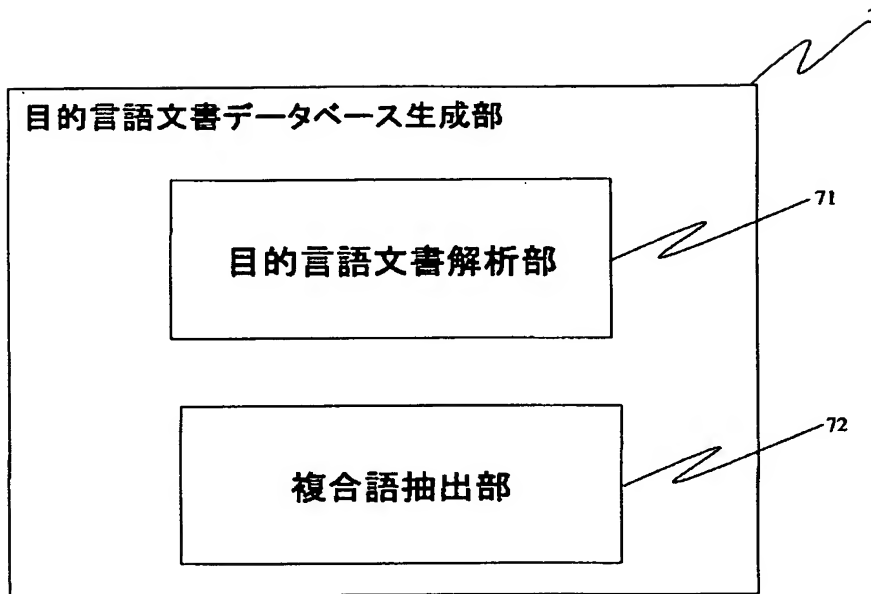
【図 2】



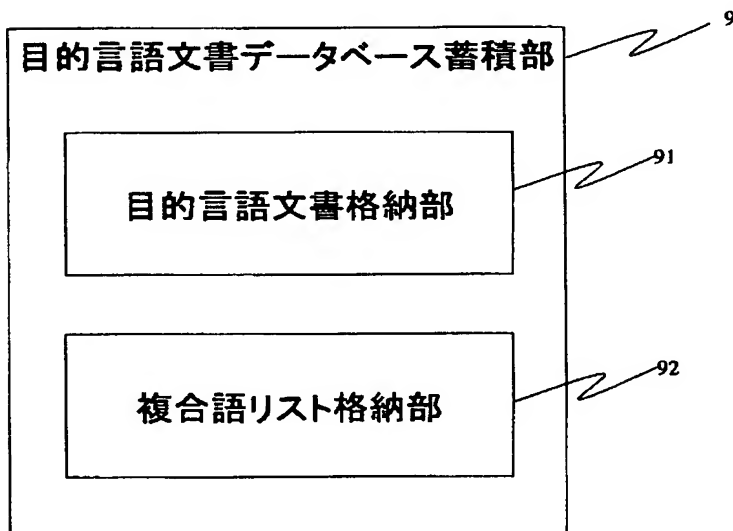
【図 3】



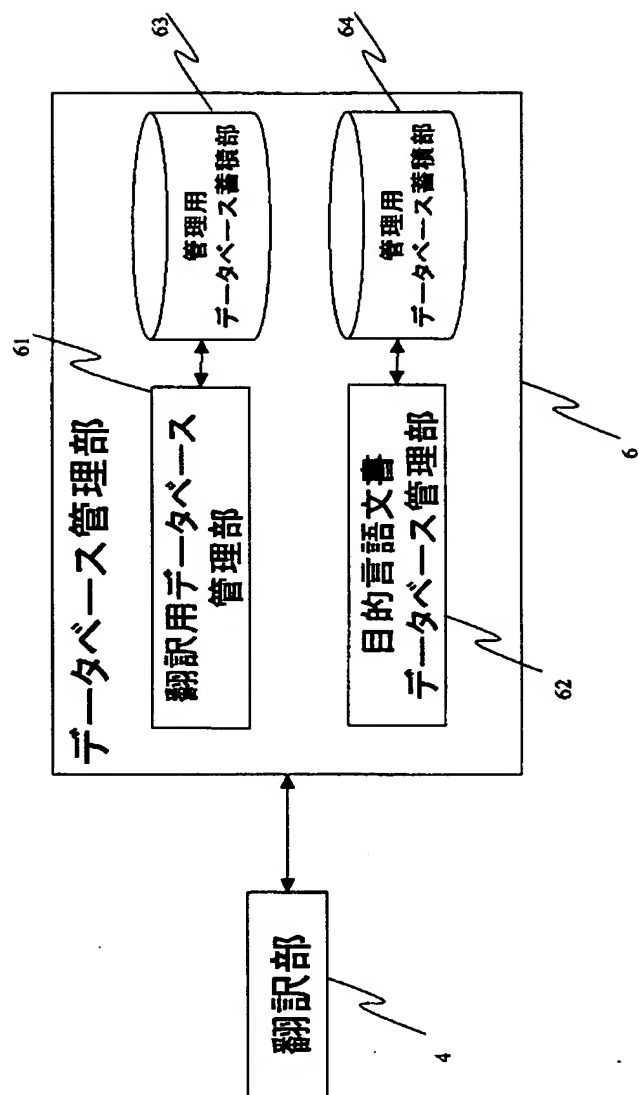
【図 4】



【図 5】



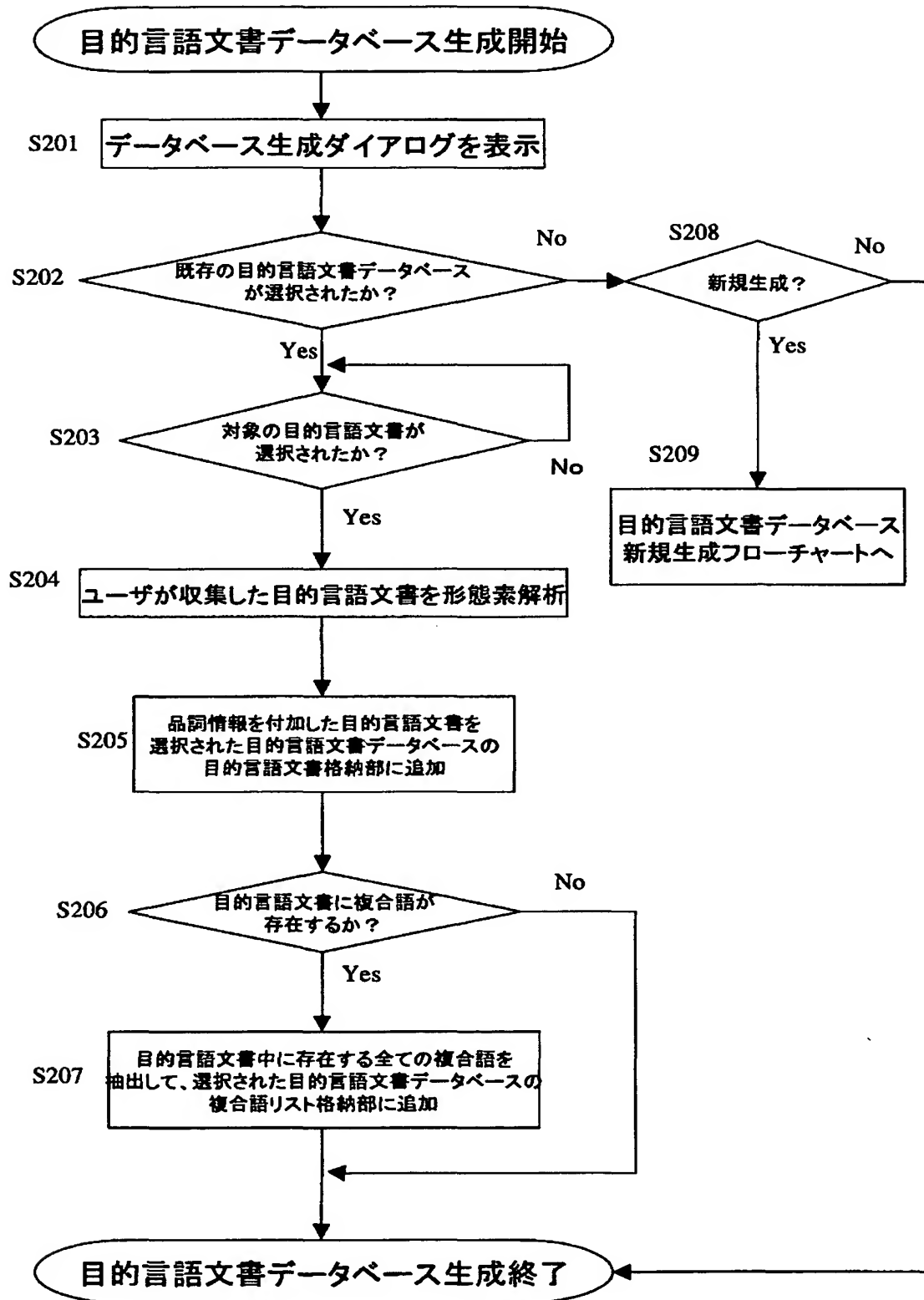
【図 6】



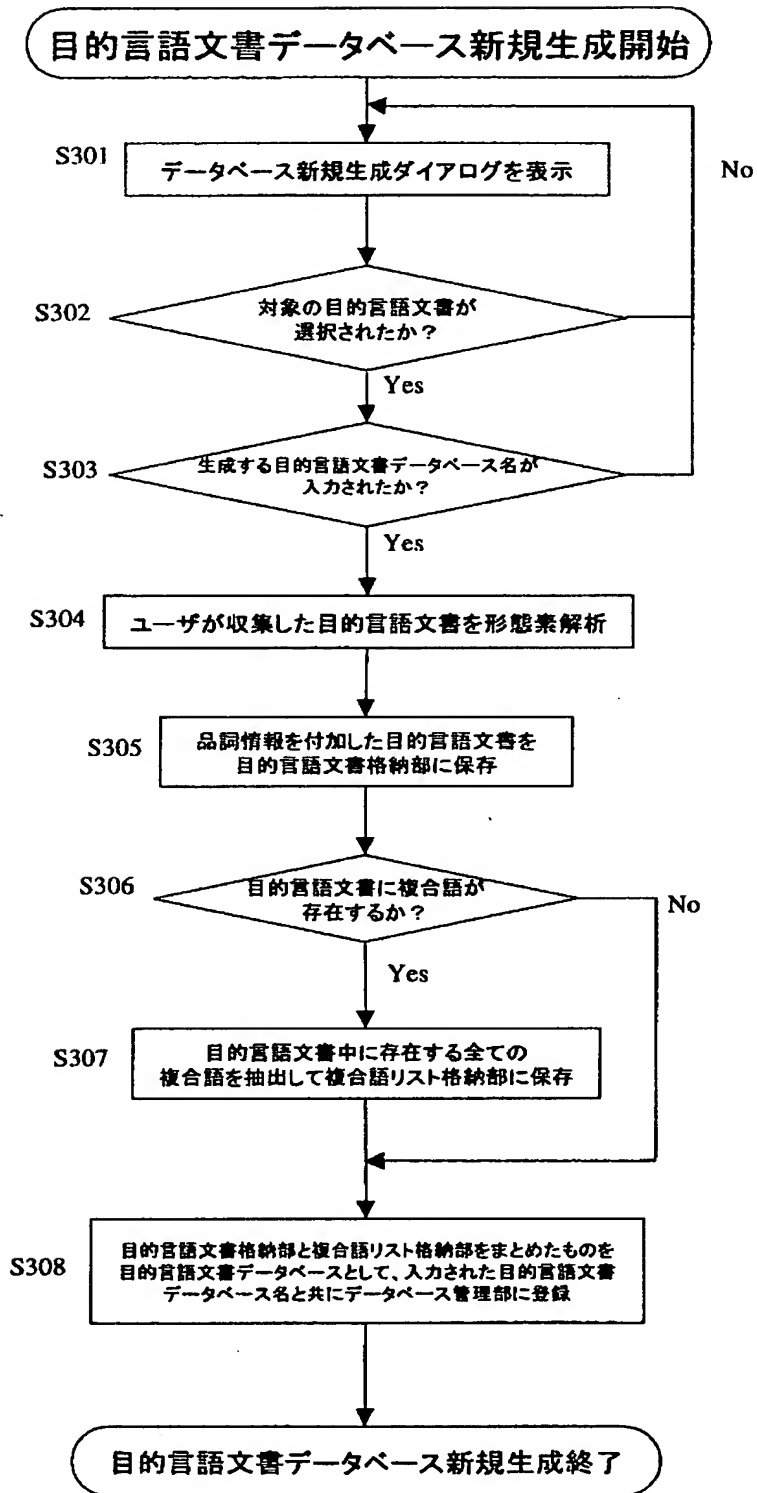
【図 7】

データベース名	優先度	訳語選択 不要単語	原文データ
データベースA	3	W1、W2、 W3	文書1 文書2
データベースB	2	W2、W4	文書3
データベースC	1	W5	文書4 文書5 文書6
...	...	...	...

【図 8】



【図 9】



【図 10】

データベース生成ダイアログ

目的言語文書データベース選択

データベースA	<input type="checkbox"/>
データベースB	<input type="checkbox"/>
データベースC	<input type="checkbox"/>
データベースD	<input type="checkbox"/>

目的言語文書名入力

キャンセル 新規生成 追加

【図 11】

データベース新規生成ダイアログ

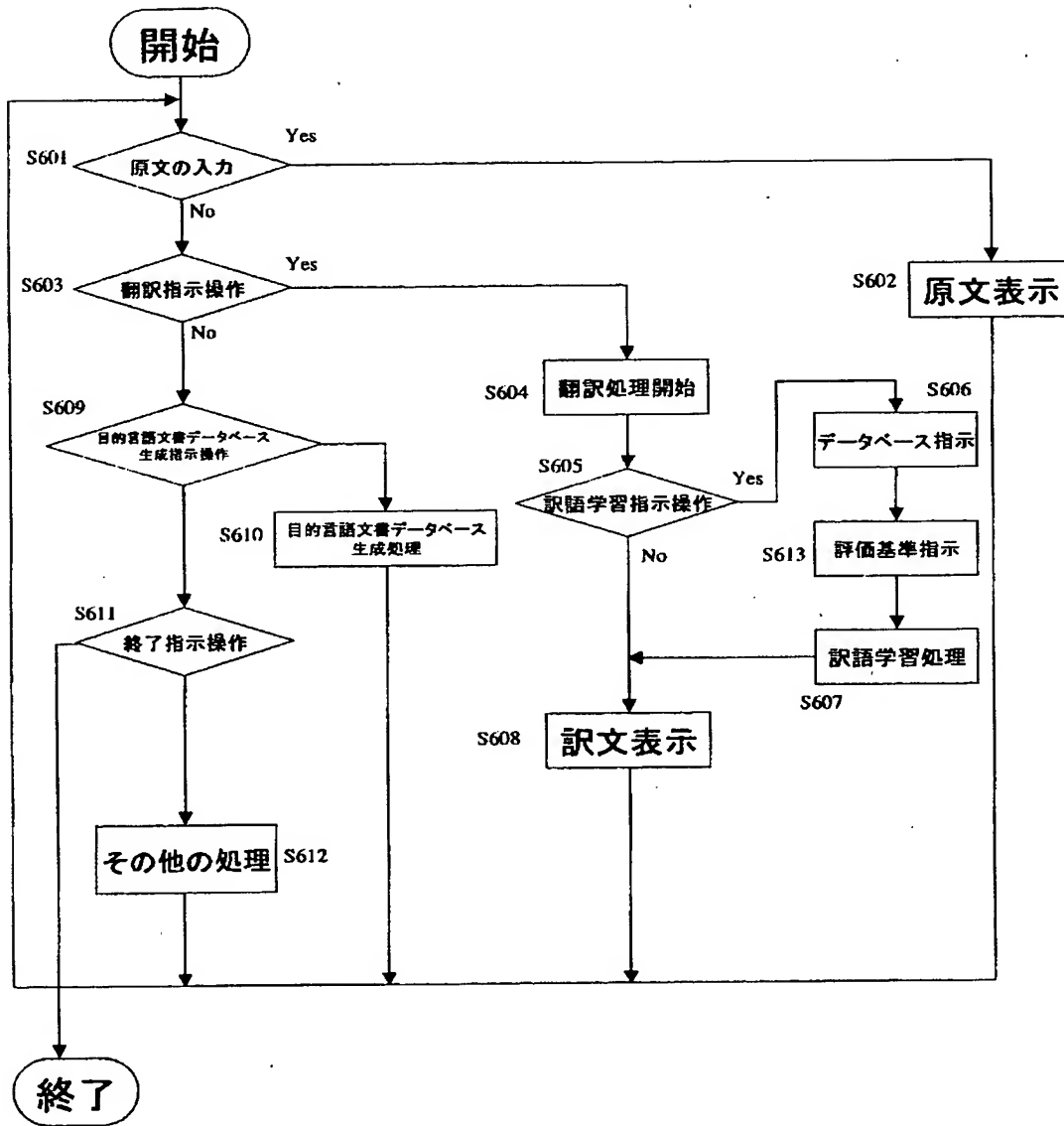
目的言語文書データベース名入力

目的言語文書名入力

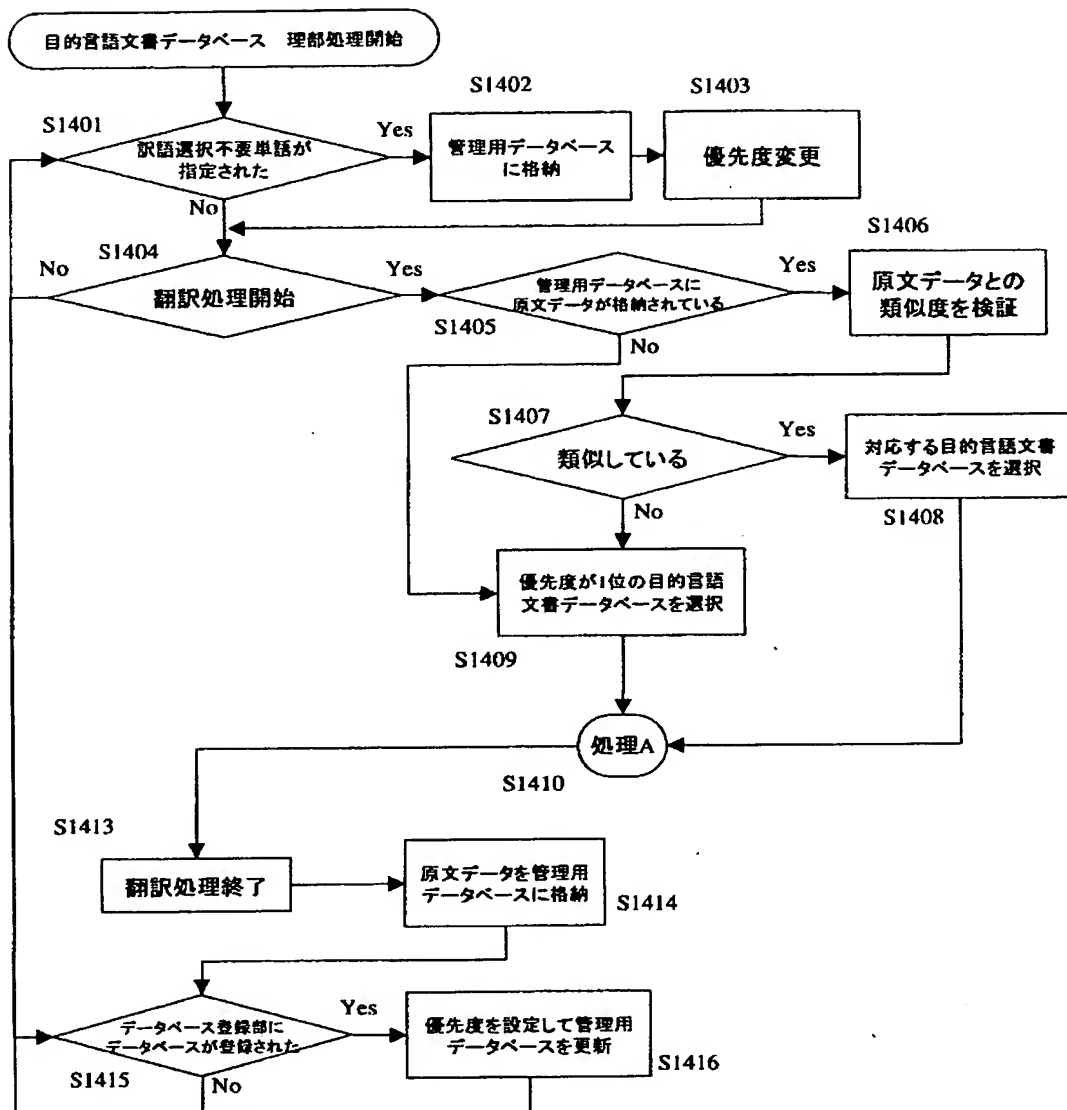
キャンセル 生成



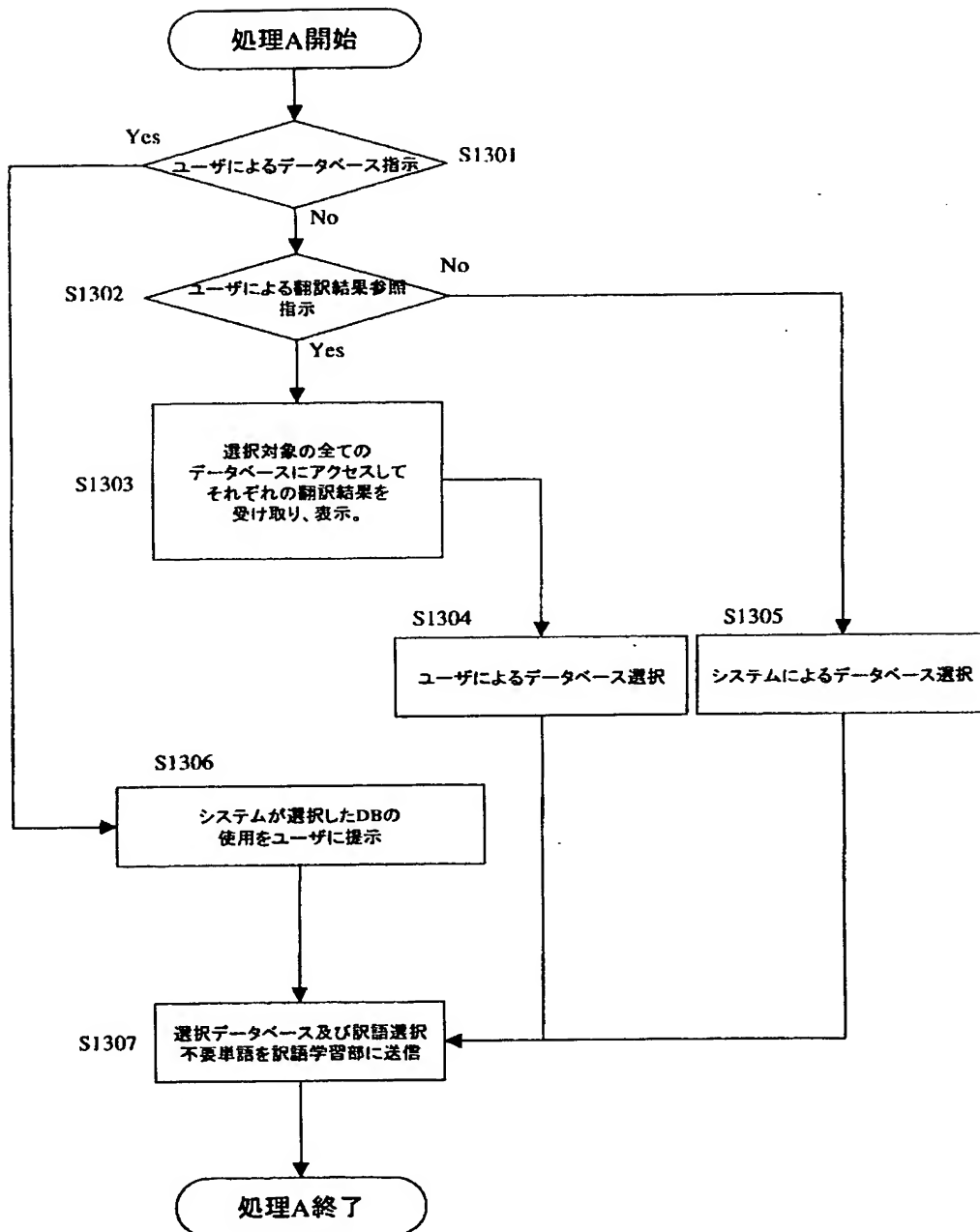
【図12】



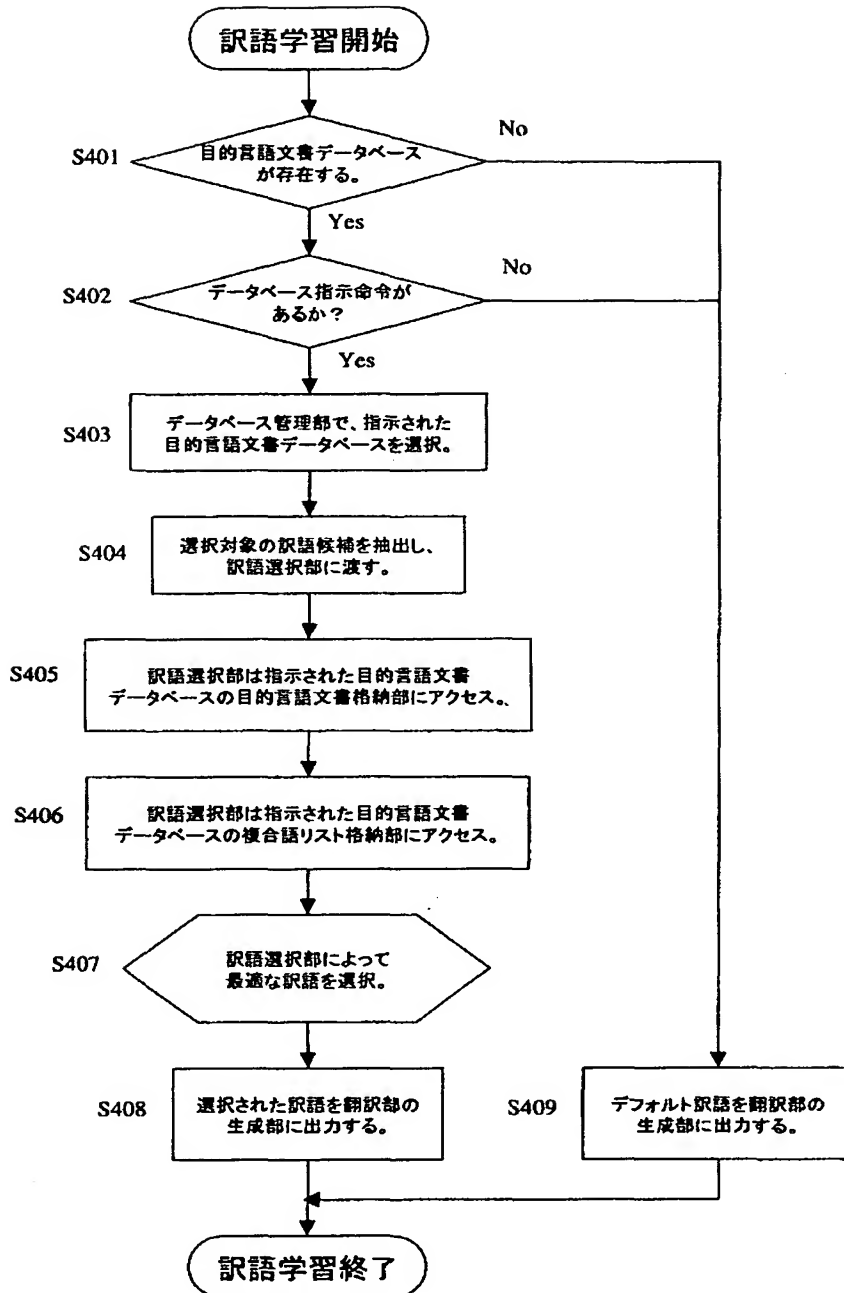
【図 13】



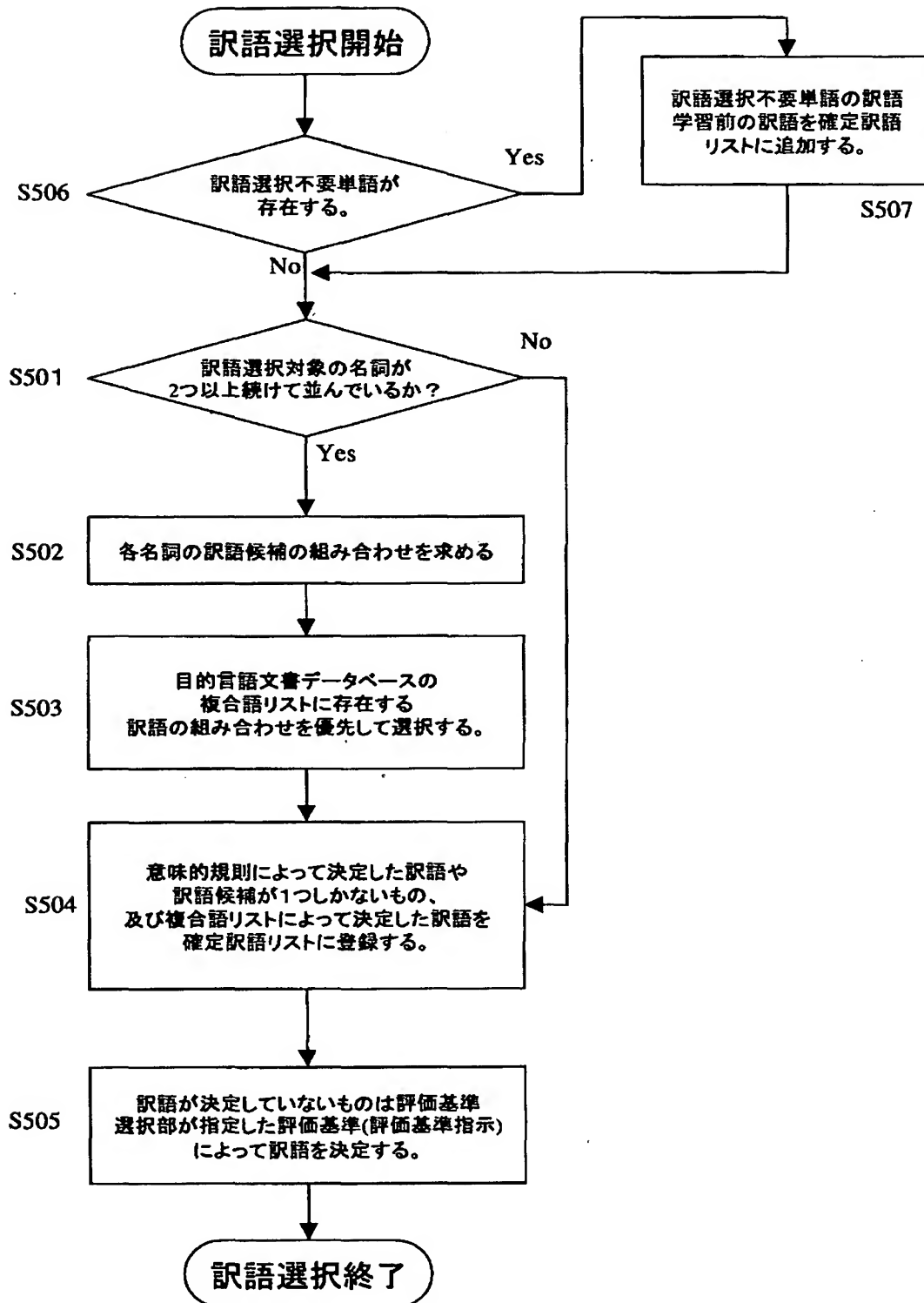
【図 1 4】



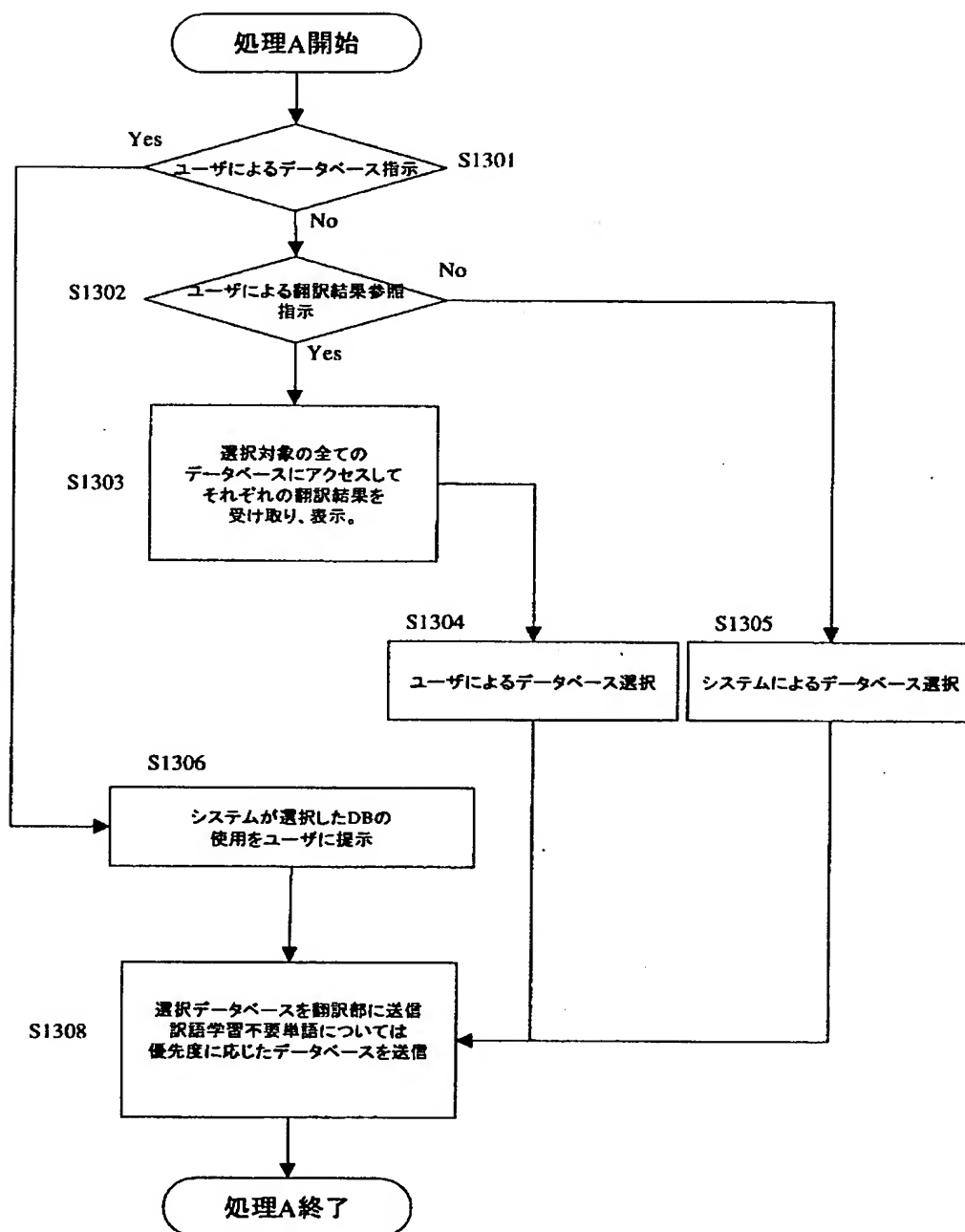
【図 1 5】



【図 16】



【図 1 7】



【書類名】 要約書

【要約】

【課題】 ユーザの嗜好、使用目的、使用分野等を反映した自動的な訳語選択を実現する。

【解決手段】 目的言語文書データベース生成部 7 は、ユーザの指示に従って、複数の目的言語文書から目的言語文書データベースを構築する。データベース管理部 6 は、1 つ以上の目的言語文書データベースを管理し、ユーザの指示に従って、各目的言語文書データベース毎に、訳語選択が不要な訳語選択不要単語を設定する。データベース管理部 6 は、例えば訳語選択不要単語数に応じて目的言語文書データベースに優先度を付し、優先度が最も高い目的言語文書データベースを翻訳部 4 に通知する。翻訳部 4 は、データベース管理部 6 によって指定された目的言語文書データベースを用いて訳語学習を行う。これにより、訳語選択はユーザの嗜好、使用目的、使用分野等を反映したものとなる。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000003078]

1. 変更年月日	2001年 7月 2日
[変更理由]	住所変更
住 所	東京都港区芝浦一丁目1番1号
氏 名	株式会社東芝